Interim Report
Tech. Publication 30
Contract FAA/BRD-363
Project 201-2

FAA

Prepared for the **FEDERAL AVIATION AGENCY**
SYSTEMS RESEARCH AND DEVELOPMENT SERVICE

DEVELOPMENT of STATISTICAL OPERATORS for PREDICTION of LOW CLOUDS

JUNE 1963

7044-72

THE TRAVELERS RESEARCH CENTER INC.

Interim Report
Technical Publication 30
Contract FAA/BRD-363

DEVELOPMENT OF STATISTICAL OPERATORS FOR PREDICTION OF LOW CLOUDS

Duane S. Cooley
Abraham M. Pavlowitz

June 1963

Project 204-2

This report has been prepared by The Travelers Research Center, Inc., for the Systems Research and Development Service, Federal Aviation Agency, under Contract FAA/BRD-363. The contents of this report reflect the views of the contractor, who is responsible for the facts and accuracy of the data presented herein, and do not necessarily reflect the official views or policy of the FAA.

THE TRAVELERS RESEARCH CENTER, INC.
650 Main Street Hartford 3, Connecticut

The Travelers Research Center, Inc.
650 Main Street, Hartford 3, Connecticut.
DEVELOPMENT OF STATISTICAL OPERATORS FOR PREDICTION OF LOW CLOUDS,
Duane S. Cooley and Abraham M. Pavlowitz. June 1963.
24 pp. incl. 3 illus., 8 tables, 13 refs.; interim rpt.
(FAA/BRD-363)

ABSTRACT

A study was made of the statistical prediction of low-cloud amounts and cloud-base heights. Cloud data and other atmospheric parameters over the central and eastern United States were analyzed on a grid mesh of approximately 52 mi (1/4-NWP grid). Predictability of low-cloud amount was evaluated by using the screening-regression method and testing the significance of the selected predictors. Predictors considered were low-cloud amount, empirically normalized cloud height, pressure, 850-mb height, surface and 850-mb temperature and dew-point spread, 850-mb geostrophic wind, and derived terms such as vorticity, divergence, and advection. The regression equations were tested on independent data. The equations may be useful for short-period prediction because they provide a better cloud forecast than persistence. They would probably be improved by including other predictors and by extending the area from which the predictors are chosen.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1.0 INTRODUCTION

Despite the meteorologist's present knowledge of the prediction of large-scale free-atmospheric flow and the nature of the microscale physics of clouds, he is not yet able to adequately express the physics of meso-scale-to-large-scale formation and dissipation of cloudiness in the form of mathematical relationships between the changes in cloudiness and the routinely observed weather parameters. Reid [12], in a study of mathematical expressions relating to cloud formation and change in ceiling height, concludes that the problem of ceiling and cloud prediction should be approached through a statistical method in which predictors are selected on the basis of physical reasoning from diagnostic equations such as those he derived. The present effort deals with the development of empirical equations for predicting cloudiness from initial values of parameters believed to have physically significant relationships to the cloudiness. This report describes the development and testing of equations for the prediction of low clouds for periods of 3, 6, 9 and 12 hr, from surface and lower-atmospheric parameters.

The occurrence of low cloudiness is of great importance in aircraft operations, and a rapid method for objective forecasting of low cloudiness is essential. The objective procedure requires automatic processing and analysis of input meteorological data in preparation for the use of such an objective cloud-forecasting technique in the Common Aviation Weather System (CAWS). This report shows the results of using a particular statistical method for prediction of low clouds and the associated data-handling and analysis procedures.

The developmental test was based on data for selected hours in October, November, and December 1962 over the United States from the Rocky Mountains eastward. Independent verification is based on data from 16 Jan. 1963 to 1 Feb. 1963. The particular hours selected were partly a function of success achieved in gathering data automatically from the Automatic Data Interchange System (ADIS) Service A airways weather-data drop and processing them on the IBM 7090 computer at the National Aviation Facilities Experimental Center (NAFEC) of the Federal Aviation Agency at Atlantic City, N. J.

Relationships were derived in the form of generalized operators---that is, statistical cloud-prediction equations applicable to the complete region rather than restricted to individual points. A study was made of cloud conditions in terms of the amount of clouds with bases in the layer below 6800 ft above the surface. The generalized operators for prediction were produced from gridpoint data based on analyzed fields of the cloud parameters and predictor parameters. The screening-regression method was used to relate physically meaningful meteorological quantities to the predicted cloud field.

This report is primarily a description of the development and testing of a statistical procedure for cloud prediction, but it also includes a brief summary of the analysis and other processing methods required to prepare data for obtaining the statistical forecasting operators.

2.0 PROCESSING AND ANALYSIS OF DATA

2.1 Analysis Methods

The procedure used for the analysis of cloud data and other meteorological data for this problem was the successive-approximation technique (SAT), a method similar to that in use at the Numerical Weather Prediction (NWP) Unit of the National Meteorological Center (NMC). This technique was outlined by Cressman [7], and specifications for the cloud-analysis program at TRC were written by Aubert [2]. A detailed description and evaluation of the cloud-analysis method is given by Davis [8]. Other parameters were analyzed by a different version of the SAT program, designed by Thomasell and Welsh [13].

From station observations, the SAT program produces parameter values at points on the NWP grid or at points located at lesser intervals along this grid. SAT computes interpolated values of the variable at the grid intersections by a series of approximations to the true field. Initial-guess values are provided at each gridpoint, and the successive approximations consist of successive corrections to the gridpoint values.

Corrections computed from a single observation are limited to those gridpoints lying within a given radius of the observation. Where these radii overlap, corrections for the gridpoints are accumulated. Successive approximations to the analysis are made by repeating the correction procedures up to a maximum of seven times, usually with stepwise decreases of the influence radius. A smoothing operator may be applied to the analysis approximations.

2.2 Data

The time periods for which data were used in this study are shown in Table 2-1. Both surface and 850-mb data over the eastern United States were used. The surface and upper-air data were cycled every 12 hr (00, 12, 00Z, etc). The cloud data were cycled every 3 hr (00, 03, 06Z, etc). On the average, about 300 stations contributed surface and cloud observations, and about 58 stations contributed the upper-air observations.

2.3 Grid

The grid used in this study is a 37 × 29 array over the central and eastern United States, with gridpoint intervals equal to 1/4 of the NWP interval. The lower-left and upper-right grid coordinates for this array are $(i_{min}, j_{min}) = (84, 29)$ and $(i_{max}, j_{max}) = (120, 57)$, respectively (see Fig. 2-1). The standard longitude of this grid was rotated 22° westward from that of the NWP grid so that $i = 92$ coincides with longitude 102°W.

TABLE 2-1
TIME PERIODS OF DATA USED IN THIS STUDY

(a) Dependent data, 1962

From	through
Oct 18, 12Z	Oct 23, 00Z
Oct 24, 12Z	Oct 27, 12Z
Oct 28, 12Z	Oct 30, 12Z
Nov 20, 00Z	Nov 21, 00Z
Nov 24, 12Z	Nov 27, 00Z
Nov 28, 12Z	Nov 30, 12Z
Dec 2, 12Z	Dec 5, 12Z

(b) Independent data, 1963

From	through
Jan 16, 00Z	Jan 20, 00Z
Jan 21, 00Z	Jan 26, 12Z
Jan 27, 12Z	Feb 1, 12Z

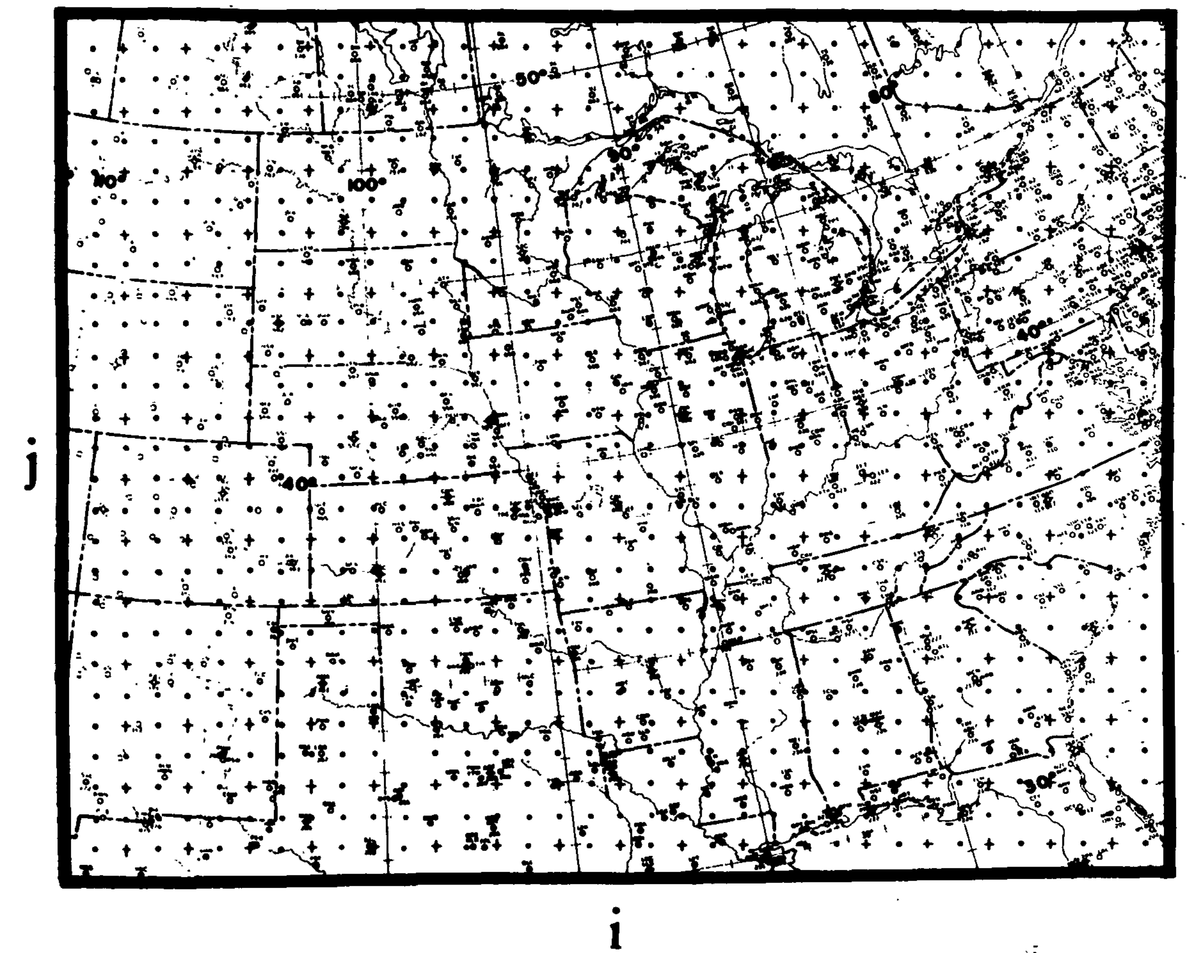


Fig. 2-1. Map of analyzed area, showing the rotated 1/4-NWP grid mesh. (i = 92 coincides with longitude 102°W.)

2.4 Preparation and Analysis of Data

Cloud observations and surface observations of pressure, temperature, dew point, and wind for this study were available on magnetic tape. These data had been prepared from hourly airways observations by procedures previously described [8, 13]. These data were further processed through an item-separator program, and the cloud data were run through a layering program. Each parameter was then run through a preprocessor and an analysis program to produce the necessary gridpoint values. Because upper-air observations were not available on magnetic tape, station observations were tabulated for card punching. Magnetic tapes were then prepared in much the same format as for the surface-data tapes. Although the general method used for the objective analysis of cloud, surface, and upper-air data was the successive-approximation technique (SAT, see Section 2.1), different programs were used for analysis of different types of parameters, as noted above.

3.0 THE PREDICTION TECHNIQUE

Heights of cloud bases and cloud amounts below 6800 ft above the surface were analyzed, and prediction equations for these quantities were developed from statistical generalized operators. The cloud amount N_1 used here is actually the amount of sky covered* by the lowest layer of clouds with a base below 6800 ft. Cloud heights of 6800 ft or more are taken as unlimited. The height of the cloud base as used in the prediction equations is an empirically normalized height determined by the method of Bryan [4] from 3014 values of cloud height over the eastern United States in September 1960. The predictands for every 3 hr and the predictors for every 12 hr were available on magnetic tape for the sampling periods and were analyzed by SAT. The 850-mb data were tabulated by hand from NMC facsimile analyses, for card punching. Screening regression was used to relate the cloud parameters and other meteorological quantities as predictors to the predictand cloud fields. The screening method is a particular form of multiple-regression prediction.

3.1 Prediction

Consider the multiple-regression prediction problem in the matrix form

$$\hat{Y} = BF, \qquad (3\text{-}1)$$

where $\hat{Y}$ is the $m' \times n$ matrix of predictand time series, B is the statistical forecasting operator ($m' \times m$ matrix of regression coefficients), and F is the $m \times n$ matrix of predictor time series. The number of predictors is m, m' is the number of predictands, and n is the number of observations in each time series. The operator B is obtained by the least-squares method, in which the sums of the squares of the forecast errors in the developmental sample are minimized.

It can be shown [1] that the matrix B can be obtained by the solution of

$$BR = A, \qquad (3\text{-}2)$$

where A is the matrix of covariances between predictors and predictands, and R is the symmetrical matrix of the covariances of the independent variables (predictors). In principle, the solution of Eq. (3-2) for the operator B may be obtained by the inversion of R:

$$B = AR^{-1}. \qquad (3\text{-}3)$$

*As input to analysis, reported cloud categories were given the following numerical values: clear, 0 tenths of sky cover; scattered, 3 tenths; broken, 7.5 tenths; and overcast, 10 tenths. The resulting objective analysis has values ranging from, say, 0.0000 tenths to 10.0000 tenths.

If some of the predictors are highly correlated in time, the matrix R may be nearly singular, in which case its inverse may be difficult if not impossible to obtain on a computing machine. If one attempts to use a large number of multiple time series of closely spaced meteorological parameters as predictors, one frequently finds that high correlations among the predictors lead to nearly singular matrices.

If the number of predictors is large, the reduction in variance obtained in the developmental (dependent) sample cannot be expected to be maintained when the operator B in Eq. (3-1) is applied to an independent sample. Lorenz [10] has shown essentially that

$$S'' \approx S' - \frac{m(n+1) + m(n-1)}{(n-1)(n+1)} R_0 \approx S' - \frac{2mR_0}{n}, \tag{3-4}$$

where S'' is the expected reduction in variance of an independent sample with the application of a statistical operator for which S' is the reduction of variance within the dependent sample, and R_0 is the ratio of the unexplained variance to the total variance in the population from which the samples were drawn. Here, both samples are assumed to consist of n observations of each of m predictors. Thus, a reduction in the number of independent variables by some process, such as selection or screening of predictors, is necessary for stability of the forecasting operator.

Statistical prediction equations meeting the above requirement were developed by the method of screening regression described by Miller [11] and based on a paper by Bryan [3]. The method deals with a predictand variable and a large set of predictor variables, selects a significant subset of predictor variables, and relates the variables by a linear multiple-regression equation. If any two possible predictors are very highly correlated with each other, one of them may be eliminated by the program.

A predictand Y is equated to a linear function of a number of predictors X_i ($i = 1, 2, \ldots, m$) [elements of the matrix F in Eq. (3-1)], where the multiple-regression coefficients b_i are obtained by the method of least squares:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_mX_m. \tag{3-5}$$

If conventional multiple-regression analysis is performed on a large number of possible predictors, not all the coefficients b_i may prove significant. Elimination of the insignificant coefficients requires considerable calculation, and this usually results in modification of the other coefficients.

The screening multiple-regression method suggested by Bryan involves a forward procedure of acceptance of predictors. This forward approach selects predictors in a stepwise manner. The variances (var) of the predictand and all predictors, and all covariances (cov) between all variables are calculated first. From the covariance between the predictand and each predictor, the square of the simple linear correlation coefficient is computed:

$$r^2(Y, X_i) = \frac{[\mathrm{cov}\ (Y, X_i)]^2}{\mathrm{var}\ (Y)\ \mathrm{var}\ (X_i)}. \tag{3-6}$$

The first selected predictor, X_j, must satisfy

$$r^2(Y, X_j) > r^2(Y, X_i) \qquad (i, j = 1, 2, \ldots, m; \quad i \neq j). \tag{3-7}$$

An F-test [1] tests the significance of this predictor. If X_j is significant, the predictand and all other predictors are orthogonalized with respect to X_j, and new covariances of the orthogonalized Y with respect to the remaining orthogonalized X_i are computed. The correlations $r(Y, X_i)$ are again calculated, and their squares are compared. The next-best predictor is selected, and its significance is tested. The process may continue until a selected predictor fails to pass the significance test or until an arbitrary number of predictors has been selected.

3.2 Preparation and Selection of Data for Generalized Operator

The variables chosen as possible predictors to be tried in this study included available observed parameters (as analyzed on the grid) and gridpoint values of derived parameters believed to be valuable for cloud prediction, such as dew-point spread, static stability, vorticity (as measured geostrophically by the Laplacian of pressure or height), wind divergence, and advection of temperature, moisture, and vorticity. Where derived predictors were required, parameter difference, gradient or Laplacian of parameter, and other numerical operations were performed by a grid-arithmetic routine [5]. Table 3-1 lists the predictors that were prepared.

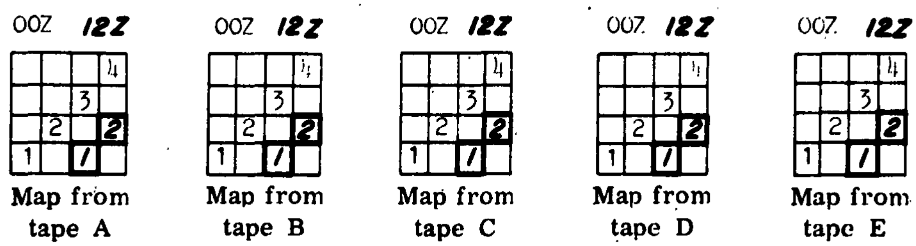
A program [6] was written that selects both the predictand value at a given grid location (i, j) and an A × B subset of predictors around the predictand. The user specifies both the number of predictand points per hour and the spatial relationship between the number of predictand points and the subset of predictors. All predictand points were over land (except those over the Great Lakes) and on any one map were separated by at least five grid intervals for increased spatial independence of cases. Predictor subset areas on each map were not allowed to overlap. Data-selection runs used 36 predictor maps 12 hr or more apart with 10 predictand points per map. Different subset areas were used on successive maps.

A simplified description of the program is given below. Figure 3-1 is a schematic representation of the data selection.

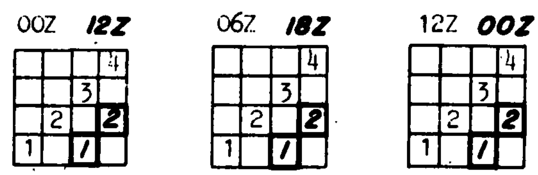
TABLE 3-1
TYPES OF PREDICTORS USED IN PREDICTING LOW-CLOUD AMOUNT AND HEIGHT BY SCREENING REGRESSION

Symbol	Unit of measurement	Definition
N_1	0.1 sky covered	Amount of lowest cloud below 6800 ft
H_1'	Dimensionless	Cloud heights transformed to a nearly normal distribution
P	mb	Sea-level pressure
T	°F	Temperature at surface
$T - T_d$	°F	Dew-point spread at surface
85T	°C	Temperature at 850 mb
$85(T - T_d)$	°C	Dew-point spread at 850 mb
85T - SFCT	°C	Measure of stability
85Z - 100Z	10 ft	850-1000-mb thickness
$-V\cdot\nabla T$	knot °F ft^{-1}	Advection* of surface temperature
$-V\cdot\nabla(T - T_d)$	knot °F ft^{-1}	Advection* of surface dew-point spread
$\nabla^2 P$	mb ft^{-2}	Laplacian of surface pressure
$-V\cdot\nabla(\nabla^2 P)$	knot mb ft^{-3}	Advection* of Laplacian of surface pressure
$\nabla\cdot V$	sec^{-1}	Surface wind* divergence
$-85V\cdot\nabla(T - T_d)$	10 °C sec^{-1}	Geostrophic advection* of 850-mb dew-point spread
$85(u_0)$, $85(v_0)$	10 ft sec^{-1}	850-mb geostrophic grid wind components
85Z	10 ft	Height of 850-mb surface
$85\nabla^2 Z$	10 ft^{-1}	Laplacian of 850-mb height

*Throughout this report, V represents the horizontal wind vector.



(a) Predictors. Numbered squares represent subset areas.



(b) Maps of clouds. Numbered squares represent subset areas.

Record 1. I_1, J_1, $(C_0)_1$, $(C_6)_1$, $(C_{12})_1$, $(P_0)_1$, $(P_0)_1$, $(P_0)_1$, $(P_0)_1$, $(P_0)_1$.

Record 2. I_2, J_2, $(C_0)_2$, $(C_6)_2$, $(C_{12})_2$, $(P_0)_2$, $(P_0)_2$, $(P_0)_2$, $(P_0)_2$, $(P_0)_2$.

Record 3. I_3, J_3, $(C_0)_3$, $(C_6)_3$, $(C_{12})_3$, $(P_0)_3$, $(P_0)_3$, $(P_0)_3$, $(P_0)_3$, $(P_0)_3$.

Record 4. I_4, J_4, $(C_0)_4$, $(C_6)_4$, $(C_{12})_4$, $(P_0)_4$, $(P_0)_4$, $(P_0)_4$, $(P_0)_4$, $(P_0)_4$.

(c) Output records. Each line is one record, including A × B values for each type of predictor and the predictands. Predictors for the first set of predictors have all been used. Read in the next set of predictors and the next cloud map. Delete the first cloud map. Generate output.

Record 5. I_1, J_1, $(C_{12})_1$, $(C_{18})_1$, $(C_0)_1$, $(P_{12})_1$, $(P_{12})_1$, $(P_{12})_1$, $(P_{12})_1$, $(P_{12})_1$.

Record 6. I_2, J_2, $(C_{12})_2$, $(C_{18})_2$, $(C_0)_2$, $(P_{12})_2$, $(P_{12})_2$, $(P_{12})_2$, $(P_{12})_2$, $(P_{12})_2$.

(d) Output records. Each line is one record, including A × B values for each type of predictor and the predictands. Predictands for the second set of predictors have all been used. Read in the next set of predictors and the next cloud map. Generate output. Continue until data are exhausted.

Fig. 3-1. Formation of records by the data-selection routine.

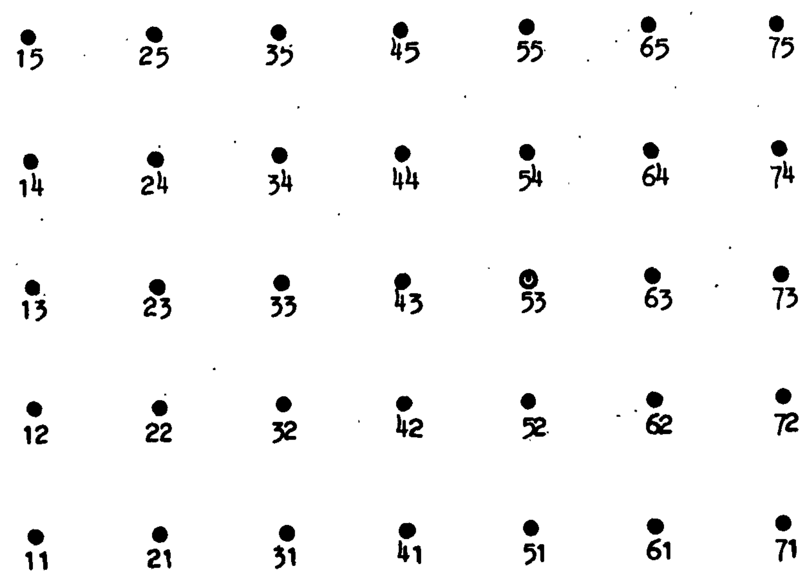


Fig. 3-2. Relative (i, j)-locations of predictor points and predictand point (53, circled).

1. Read in N cloud maps (e.g., 00, 06, and 12Z).*
2. Read in M predictor maps for 00Z (e.g., pressure, temperature, and dew point).
3. Store the first predictand cloud value for each of the N cloud maps.
4. Compute the subset area corresponding to the first predictand point for each map.
5. Store each field of predictors for the predictand point.
6. Repeat steps 3, 4, and 5 until all the predictand points on the N maps are used.
7. Read in the next set of predictor maps (e.g., 12Z).
8. Read in the cloud maps (e.g., 12, 18, and 00Z).*
9. Repeat steps 3 through 8 until the data are exhausted.

In the data selection for the generalized operators in this developmental test, the A $\times$ B predictor subset was taken as a conventionally oriented 7 $\times$ 5 subset, in which the relative position of the predictand point was taken as point ($i=5$, $j=3$), as shown in Fig. 3-2.

3.3 Development of Prediction Equations

A screening-regression program written by Enger and Rodante [9] was used to derive the multiple linear-regression equations. The program consists of two parts: covariance-matrix generation and screening regression.

For the purpose of this study, the program was allowed to select 30 predictors. These predictors were subjected to an ordinary F-test of significance, and the first few that passed at the 1% level were retained. (A predictor value is taken at the predictand gridpoint and at certain of the 34 surrounding gridpoints.)

The data-selection routine prepares a tape of lagged predictand values and corresponding predictor values for input to the covariance-matrix generation routine. By generating a large covariance matrix comprising all predictands and predictors, the regression program permits the simultaneous selection at several lags of one or several predictands and their predictors.

One hundred eighty predictor and predictand variables can be selected for covariance-matrix generation. The regression program can accommodate up to 175 predictor values selected from certain of the 35 gridpoints and from different types of predictors in the data subset, in addition to the predictand value at each of five lags. In developing these prediction equations, 140 predictor variables were used for each lag. A set of screening-regression runs relates certain types of predictors to the predictand at the four lags (3, 6, 9 and 12 hr), with a separate run for each lag. Zero-lag covariance calculations are included as a means of checking correctness of selected data. This study included the variables shown in Table 3-1

*These predictand hours were cited for the purpose of illustration. Actual predictand times were 00, 03, 06, 09, and 12Z for a 00Z predictor map and 12, 15, 18, 21, and 00Z for a 12Z predictor map.

TABLE 3-2
TYPES, NUMBERS, AND RELATIVE LOCATIONS OF POSSIBLE PREDICTORS FOR FORECASTING LOW-CLOUD AMOUNT AND HEIGHT

Symbols*	Number of gridpoints	Relative (i,j)-locations of predictor gridpoints†
N_1, H_1'	21	11 31 51 71 42 52 62 13 23 33 43 53 63 73 44 54 64 15 35 55 75
$T - T_d$, $85(T - T_d)$, $85Z - 100Z$, $\nabla^2 P$, $85\nabla^2 Z$	10	11 31 51 71 33 53 73 15 55 75
P, $-V \cdot \nabla(T - T_d)$, $-85V \cdot \nabla(T - T_d)$	8	11 51 71 33 53 15 55 75
T, $85T$, $85T$ - SFCT, $-V \cdot \nabla T$, $-V \cdot \nabla(\nabla^2 P)$, $\nabla \cdot V$, $85(u_0)$, $85(v_0)$, $85Z$	5	11 71 53 15 75

*Symbols are defined in Table 3-1.
†See Fig. 3-3.

as possible predictors for the low-cloud-amount predictand $\hat{N}_1$ and the empirically normalized low-cloud-height predictand $\hat{H}_1'$. Table 3-2 lists the number and relative locations of grid points at which each possible type of predictor was selected as input to the screening program. (The 21 transformed heights were not used as possible predictors of cloud amounts, nor were the 21 values of cloud amount used as possible predictors of height.) Table 3-3 gives the transformed values of height categories.

TABLE 3-3
EMPIRICALLY NORMALIZED
VALUES OF CLOUD-BASE HEIGHT

Height H, ft	H_1'
$0 \le H < 400$	-2.550
$400 \le H < 1000$	-1.840
$1000 \le H < 2000$	-1.280
$2000 \le H < 3300$	-0.809
$3300 \le H < 5000$	-0.414
$5000 \le H < 6800$	-0.233
$6800 \le H$	+0.710

4.0 RESULTS

Each screening run was allowed to select as many as 30 predictors for each time lag of the predictand so that the significance of many possible predictors might be examined. The screening program thus selected an arbitrary number of predictors but calculated and printed a value of F [1] with each predictor to permit testing of the significance of the predictor.

Regression equations were produced for time lags of 3, 6, 9, and 12 hr for the predictands of low-cloud amount and transformed low-cloud height. The selection of 30 predictors per set of screening runs resulted in 240 equations. Only the equations containing the most significant predictors are given below. The criterion for the selection of predictors was the ordinary F-test at the 1% level, where $F_{0.01}$ has 1 and N - n degrees of freedom. For the purpose of this developmental test, N is taken as the number of cases in the sample (360), and n is the total number of predictors selected—including the predictor whose significance is being tested. Table 4-1 lists the regression equations produced for statistical prediction of low-cloud amount and height over the eastern and central United States.

The results on dependent data for the first few significant predictors of each run are presented in Table 4-2. The last column of Table 4-2 gives, for comparison, the persistence values of the percent reduction in variance of the given predictand for 3-, 6-, 9-, and 12-hr lags in the sample tested.

Although most of the predictors chosen, as seen in either Table 4-1 or 4-2, are cloud amounts or heights themselves, in each type of forecast at least one significant predictor is chosen from either the surface or the 850-mb data. It is interesting also that among the most significant surface and 850-mb predictors, the "raw" observed variables were not chosen, but rather the derived variables such as divergence, Laplacian of pressure or height, dew-point spread, and the advection of parameters that would be expected to be related to cloud development or dissipation. The results of the screening forecasts on the dependent sample given in Table 4-2 show appreciable improvement over persistence.

A method of Bryan [1] has been used for empirical normalization of the observed frequency distribution of cloud-base height, first to obtain a parameter that is defined even though no clouds may be present below 6800 ft, and second to avoid some of the difficulty that abnormality of the distribution of the predictand might introduce into prediction by a regression method. For the latter reason, cloud amount presumably also should have been normalized, although it was not in this study.

The results of screening regression indicate that the best predictors of clouds for 3-, 6-, 9-, and 12-hr periods are low-cloud amounts at initial time. For a 3-hr prediction, the surface-wind divergence and the 850-mb dew-point spread were also selected. The Laplacian of the surface pressure (measure of relative geostrophic vorticity at sea level) was selected as a 6-hr predictor. The v-component of the

TABLE 4-1
REGRESSION EQUATIONS FOR FORECASTING LOW-CLOUD AMOUNT AND NORMALIZED CLOUD HEIGHT

(a) Prediction equations* for cloud amount, $(\hat{N}_1)_{53}$

Time lag, hr	Equation
3	$(\hat{N}_1)_{53} = 0.98804 + 0.49109(N_1)_{53} + 0.16091(N_1)_{44} + 0.13922(N_1)_{42} - 20411.0\ \nabla\cdot V_{75} - 0.062103\ [85(T - T_d)]_{51}$
6	$(\hat{N}_1)_{53} = 0.45203 + 0.19787(N_1)_{42} + 0.26521(N_1)_{54} + 0.18216(N_1)_{13} + 1.8758 \times 10^{10}\ \nabla^2 P_{13} + 0.13496(N_1)_{31}$
9	$(\hat{N}_1)_{53} = 0.64434 + 0.24391(N_1)_{53} + 0.17065(N_1)_{35} + 0.19249(N_1)_{31} + 0.093543\ [85(v_0)]_{53} + 0.19255(N_1)_{54}$ $+ 7020.1\ (-V\cdot\nabla T)_{53} - 7883.0\ [-V\cdot\nabla(T - T_d)]_{75}$
12	$(\hat{N}_1)_{53} = 2.2528 + 0.23716(N_1)_{44} + 0.13591(N_1)_{31} + 0.14148\ [85(v_0)]_{75} + 0.21295(N_1)_{35} - 0.092122\ [85(T - T_d)]_{71}$ $+ 9666.3\ (-V\cdot\nabla T)_{71} + 1.6849 \times 10^{15}\ [-V\cdot\nabla(\nabla^2 P)]_{75}$

(b) Prediction equations* for cloud height, $(\hat{H}_1')_{53}$

Time lag, hr	Equation
3	$(\hat{H}_1')_{53} = 0.0079173 + 0.37274(H_1')_{53} + 0.18285(H_1')_{23} + 0.18434(H_1')_{44} + 0.14090(H_1')_{51} - 2.1317 \times 10^{9}\ (85\nabla^2 Z)_{15}$ $+ 5460.6\ [-85V\cdot\nabla(T - T_d)]_{55}$
6	$(\hat{H}_1')_{53} = 4.9348 + 0.34822(H_1')_{44} + 0.17070(H_1')_{42} + 0.16893(H_1')_{15} - 0.011498(85Z - 100Z)_{71} + 0.015797\ (T - T_d)_{55}$ $+ 0.13505(H_1')_{71}$
9	$(\hat{H}_1')_{53} = -0.0013518 + 0.24568(H_1')_{44} + 0.17955(H_1')_{15} + 0.15697(H_1')_{63} - 0.034484\ [85(v_0)]_{53} + 0.15925(H_1')_{52}$ $- 0.63071000 \times 10^{10}\ \nabla^2 P_{75}$
12	$(\hat{H}_1')_{53} = -0.11163 + 0.28257(H_1')_{44} + 0.25293(H_1')_{63} - 0.039977\ [85(v_0)]_{75} - 2490.0\ (-V\cdot\nabla T)_{71} + 8662.3\ [-85V\cdot\nabla(T - T_d)]_{15}$

*Symbols are defined in Table 3-1. Numerical subscripts refer to relative grid position (i,j) of the predictor or predictand in the 7 × 5 subset of possible predictor points about the predictand point (5,3).

TABLE 4-2

RESULTS OF SCREENING REGRESSION FOR PREDICTION OF LOW-CLOUD AMOUNT AND NORMALIZED CLOUD HEIGHT (DEPENDENT DATA, 360 CASES)

(a) Predictand is low-cloud amount, $(\hat{N}_1)_{53}$

Time lag, hr	Predictors* in order of selection	Prediction % red of predictand	Rms error†	Std dev†	Persistence % red of predictand
3	$(N_1)_{53}$, $(N_1)_{44}$, $(N_1)_{42}$, $\nabla\cdot V_{75}$, $85\ (T - T_d)_{51}$	61.7	1.84	2.97	56.3
6	$(N_1)_{42}$, $(N_1)_{54}$, $(N_1)_{13}$, $\nabla^2 P_{13}$, $(N_1)_{31}$	47.0	2.12	2.91	30.4
9	$(N_1)_{53}$, $(N_1)_{35}$, $(N_1)_{31}$, $85(v_0)_{53}$, $(N_1)_{54}$, $-V\cdot\nabla T_{53}$, $-V\cdot\nabla(T - T_d)_{75}$	50.0	2.16	3.06	37.1
12	$(N_1)_{44}$, $(N_1)_{31}$, $85(v_0)_{75}$, $(N_1)_{35}$, $85(T - T_d)_{71}$, $-V\cdot\nabla T_{71}$, $-V\cdot\nabla(\nabla^2 P)_{75}$	36.7	2.54	3.20	20.7

(b) Predictand is normalized cloud height, $(\hat{H}_1')_{53}$

Time lag, hr	Predictors* in order of selection	Prediction % red of predictand	Rms error†	Std dev†	Persistence % red of predictand
3	$(H_1')_{53}$, $(H_1')_{23}$, $(H_1')_{44}$, $(H_1')_{51}$, $85\ \nabla^2 Z_{15}$, $-85\ V\cdot\nabla(T - T_d)_{55}$	50.8	0.53	0.75	39.8
6	$(H_1')_{44}$, $(H_1')_{42}$, $(H_1')_{15}$, $(85Z - 100Z)_{71}$, $(T - T_d)_{55}$, $(H_1')_{71}$	46.8	0.57	0.78	28.0
9	$(H_1')_{44}$, $(H_1')_{15}$, $(H_1')_{63}$, $85(v_0)_{53}$, $(H_1')_{52}$, $\nabla^2 P_{75}$	43.7	0.55	0.74	25.0
12	$(H_1')_{44}$, $(H_1')_{63}$, $85(v_0)_{75}$ $-V\cdot\nabla T_{71}$, $-85\ V\cdot\nabla(T - T_d)_{15}$	30.3	0.65	0.78	14.6

*Symbols are defined in Table 3-1. Numerical subscripts refer to relative grid position (i,j) of the predictor or predictand in the 7 × 5 subset of possible predictor points about the predictand point (5,3).

†Dimensions of root-mean-square (rms) error or standard deviation (std dev) of cloud amount are tenths of sky cover. Dimensions of rms error or standard deviation of cloud height are units comparable to number of standard deviations obtained by empirically normalizing an observed distribution of cloud heights.

TABLE 4-3

VERIFICATION OF REGRESSION EQUATIONS ON INDEPENDENT LOW-CLOUD-AMOUNT DATA*

(a) Contingency table for 3-hr forecasts

		Obs amount†				Fcst total
		9.0	6.0	2.0	0.0	
Fcst amount†	9.0	0	0	0	0	0
	6.0	16	15	8	5	44
	2.0	5	17	37	24	83
	0.0	1	2	19	141	163
Obs total		21	34	64	170	290

Number of hits = 193. Percent correct = 66.6.

Rms error of N_1 = 2.10. Std dev of N_1 = 3.21.

(b) Contingency table for 6-hr forecasts

		Obs amount†				Fcst total
		9.0	6.0	2.0	0.0	
Fcst amount†	9.0	0	0	0	0	0
	6.0	14	11	8	1	34
	2.0	6	21	39	34	100
	0.0	5	11	22	118	156
Obs total		25	43	69	153	290

Number of hits = 168. Percent correct = 57.9.

Rms error of N_1 = 2.70. Std dev of N_1 = 3.35.

(c) Contingency table for 9-hr forecasts

		Obs amount†				Fcst total
		9.0	6.0	2.0	0.0	
Fcst amount†	9.0	1	0	0	0	1
	6.0	11	10	5	6	32
	2.0	11	19	41	44	115
	0.0	2	10	24	105	142
Obs total		26	39	70	155	290

Number of hits = 157. Percent correct = 54.1.

Rms error of N_1 = 2.87. Std dev of N_1 = 3.30.

(d) Contingency table for 12-hr forecasts

		Obs amount†				Fcst total
		9.0	6.0	2.0	0.0	
Fcst amount†	9.0	0	0	0	0	0
	6.0	8	8	9	2	27
	2.0	11	19	38	59	127
	0.0	6	8	22	100	136
Obs total		25	35	69	161	290

Number of hits = 146. Percent correct = 50.3.

Rms error of N_1 = 3.18. Std dev of N_1 = 3.29.

*Forecasts of low-cloud amounts at gridpoints are compared with SAT analyses of cloud data at gridpoints.

†Lower limit only of each category is shown. Upper limit is less than the lower limit of the next-higher category. Units are tenths of sky covered by low clouds (below 6800 ft).

TABLE 4-4

VERIFICATION OF REGRESSION EQUATIONS ON INDEPENDENT CLOUD-HEIGHT DATA*

(a) Contingency table for 3-hr forecasts

Fcst height† (H; H')	Obs height† (H; H') 68; +0.71	50; -0.24	33; -0.42	20; -0.81	10; -1.28	4; -1.84	0; -2.55	Fcst total
68; +0.24	95	3	4	3	4	0	0	109
50; -0.32	29	1	7	11	5	2	2	57
33; -0.61	11	2	8	10	6	2	0	39
20; -1.04	9	1	5	16	6	5	1	43
10; -1.56	3	2	1	6	8	4	4	28
4; -2.19	1	0	0	1	4	2	2	10
0; -2.90	0	0	0	1	1	1	1	4
Obs total	148	9	25	48	34	16	10	290

Number of hits = 131. Percent correct = 45.2.

Rms error of H_1' = 0.76. Std dev of H_1' = 0.98.

(b) Contingency table for 6-hr forecasts

Fcst height† (H; H')	Obs height† (H; H') 68; +0.71	50; -0.24	33; -0.42	20; -0.81	10; -1.28	4; -1.84	0; -2.55	Fcst total
68; +0.24	98	3	12	16	9	0	1	139
50; -0.32	28	1	10	18	11	4	2	74
33; -0.61	5	1	3	12	6	4	2	33
20; -1.04	3	0	3	3	12	3	1	25
10; -1.56	1	0	1	1	3	2	2	0
4; -2.19	0	0	0	0	1	4	3	8
0; -2.90	0	0	0	0	0	1	0	1
Obs total	135	5	29	50	42	18	11	290

Number of hits = 112. Percent correct = 38.6.

Rms error of H_1' = 0.87. Std dev of H_1' = 1.00.

*Forecasts of cloud heights at gridpoints are compared with normalized cloud heights from SAT analyses of cloud data at gridpoints.

†Lower limit only of each category is shown. Upper limit is less than the lower limit of the next-higher category. H, listed first, is in hundreds of feet; H', listed second, is dimensionless.

(c) Contingency table for 9-hr forecasts

Fcst height† (H; H′)	Obs height† (H; H′): 68; +0.71	50; -0.24	33; -0.42	20; -0.81	10; -1.28	4; -1.84	0; -2.55	Fcst total
68; +0.24	75	1	10	10	2	1	1	100
50; -0.32	22	1	11	13	14	4	1	66
33; -0.61	17	0	9	16	9	4	4	59
20; -1.04	9	2	3	13	8	5	4	44
10; -1.56	3	0	1	0	6	3	3	16
4; -2.19	0	0	1	0	1	1	1	4
0; -2.90	0	0	0	0	0	1	0	1
Obs total	126	4	35	52	40	19	14	290

Number of hits = 105. Percent correct = 36.2.

Rms error of H_1' = 0.90. Std dev of H_1' = 1.01.

(d) Contingency table for 12-hr forecasts

Fcst height† (H; H′)	Obs height† (H; H′): 68; +0.71	50; -0.24	33; -0.42	20; -0.81	10; -1.28	4; -1.84	0; -2.55	Fcst total
68; +0.24	50	2	9	6	9	0	0	76
50; -0.32	61	1	12	16	8	8	2	108
33; -0.61	10	0	4	10	9	4	5	42
20; -1.04	8	2	6	9	11	8	3	47
10; -1.56	5	0	2	0	0	3	6	16
4; -2.19	1	0	0	0	0	0	0	1
0; -2.90	0	0	0	0	0	0	0	0
Obs total	135	5	33	41	37	23	16	290

Number of hits = 64. Percent correct = 22.1.

Rms error of H_1' = 0.99. Std dev of H_1' = 1.05.

†Lower limit only of each category is shown. Upper limit is less than the lower limit of the next-higher category. H, listed first, is in hundreds of feet; H′, listed second, is dimensionless.

850-mb geostrophic wind on the rotated NWP grid was selected as a 9-hr predictor. The advection of surface temperature and dew-point spread were also chosen. The 850-mb dew-point spread, v-component of the grid geostrophic wind at 850 mb, and the advection of surface temperature were selected as 12-hr predictors. Similar predictors were chosen for low-cloud height in addition to the heights themselves.

The results of testing the regression equations on independent data are shown in Tables 4-3 and 4-4. Table 4-3 shows the verification of low-cloud-amount forecasts on independent data. The verification in terms of percentage of hits in the prescribed categories appears better than might have been expected from the percent reduction of variance of cloud amount in the dependent sample. Notice, however, that overcast was forecast only once [Table 4-3(c)], even though skies overcast with low clouds were "observed" at the gridpoints 8% of the time. The 12-hr forecasts show another pecularity in that the rms error is large (0.3 cloud cover) even though the percentage of hits remains above 50.

The forecast categories of transformed height were arbitrarily chosen with limits usually midway between the normalized values of Table 3-3. The forecasts are obviously biased toward higher cloud bases, but, in practice, forecasts could be adjusted to allow for this deviation. The percentage of hits correct in the independent sample as shown in Table 4-4 ranges from 6 to 8% less than the percent reduction in variance of normalized cloud height in the dependent sample shown in Table 4-2.

5.0 SUMMARY AND CONCLUSIONS

This study has provided regression equations for gridpoint prediction of cloud amounts and categories of cloud-base height where the height of the cloud is less than 6800 ft above the surface. The results indicate that 3-to-12-hr prediction by objective methods is feasible when low-cloud amounts or heights themselves and suitable parameters from surface and 850-mb observations are used as predictors. The predictability is significantly better than persistence up to at least 12 hr. These conclusions are based on a dependent sample of 360 cases and an independent sample of 290 cases. The statistical operators are generalized to apply to any gridpoint over the eastern and central United States, given a grid mesh of 1/4 the NWP; this generalization alone would be expected to reduce the percent reduction of variance of the predictand below that which would be obtained from a comparable statistical operator designed to predict low-cloud amount at a single geographic location.

Gridpoint data used for the test were obtained from objective analyses, The cloud parameters and continuous field parameters were analyzed by the successive-approximation technique on an IBM 7090. Derived parameters were then calculated from the gridpoint values.

From the results, one can conclude that, for short periods, the observed surface and 850-mb variables, or terms derived from these variables, are significant predictors. For longer time periods, the advection of temperature, moisture, and vorticity around the predictand point are significant predictors. The regression equations may be useful for short-period prediction because they provide a better cloud prediction than persistence. Note that verification of cloud amount is in terms of tenths of cloudiness as analyzed on the grid. The scores might appear better if verified in cloud-amount categories, which are of more concern to users.

While the verification on the independent sample appears satisfactory for up-to-9-hr forecasts, there is a distinct forecast bias away from large cloud amounts and very low cloud bases in the layer below 6800 ft. Adjustment for this bias could be made in the forecast categories, however. Standard deviation of both height and amount was larger during the winter (independent) sample than during the fall (dependent) sample. Mean cloud height was lower and mean cloud amount was slightly greater during the winter than in the fall. These differences between the samples, of course, adversely affect the application of the regression equations to the independent sample.

6.0 RECOMMENDATIONS

The discontinuous nature of cloudiness can cause unrepresentative values to be computed at gridpoints regardless of the method used for interpolative analysis of cloud observations made from the earth's surface. The abnormality of the frequency distribution of observed cloud amounts may also adversely affect the prediction of cloudiness by a linear-regression method. Empirical normalization of the frequency distribution of cloud amount by Bryan's method might improve the prediction. Screening predictions of cloud amount should be obtained in this way and the results compared with those presented in this report.

Comparison should be made with a screening-regression run in which predictors are selected only at the relative point, (5, 3)—that is, the predictand point—to see how much improvement is obtained by selecting predictors over an area rather than at a single point.

Further improvement in cloud prediction might be obtained by adding such predictors as the rate of upslope motion and the sine or cosine of the local hour angle of the sun. Other predictors that should be tried for low-cloud prediction are the 850-mb wind divergence and the advection of vorticity at 850 mb. The number of gridpoints could be increased at which parameters appearing significant here could be obtained as possible predictors in a screening run. The predictor subset area should be extended for forecasts of from 6 to 12 hr. Prediction operators should also be developed for the western United States, although it can be expected that, because of the rough terrain of the West and the scarcity of data off the Pacific coast, the results may be poorer than those obtained in this study.

7.0 ACKNOWLEDGMENTS

The authors wish to acknowledge the computational support given this task by the National Aviation Facilities Experimental Center of the Federal Aviation Agency and by the United Aircraft Corporation's Computations Laboratory. Thanks go to IBM 7090 programmers Jerrold Ruck, formerly of UAC, and to Steven Morrison of UAC. The authors are grateful to other members of the Meteorological Analysis and Prediction Division of TRC who assisted in data preparation and analysis.

8.0 REFERENCES

1. Anderson, R. L., and T. A. Bancroft, Statistical Theory in Research. New York: McGraw-Hill, 1952.

2. Aubert, E. J., Objective Analysis of Clouds. Paper delivered at 190th Natl. Meeting, Am. Meteorol. Soc., New York, Jan. 24, 1961.

3. Bryan, J. G., Special Techniques in Multiple Regression. Unpubl. MS, 1944.

4. — and K. W. Veigas, Statistical Methods in Forecasting. Final Rpt. 7065-31, The Travelers Research Center, Inc., Hartford, Aug. 1962.

5. Cooley, D. S., and N. E. Bowne, Program Specifications For Arithmetic Subroutines Applied to Grid Point Data. The Travelers Research Center, Inc., Hartford, 1962.

6. Cooley, D. S., and A. M. Pavlowitz, Program Specifications for a Routine for Data Selection to be Used in the Development of a Statistical Operator. The Travelers Research Center, Inc., Hartford, 1962.

7. Cressman, G. P., "An Operational Objective Analysis System," Mo. Weath. Rev. 87:10, 367—374 (1959).

8. Davis, E. L., Objective Techniques for the Analysis of Clouds and Ceilings. Tech. Rpt. 7044-35, The Travelers Research Center, Inc., Hartford, Nov. 1962.

9. Enger, I., and F. R. Rodante, Screening Regression II, Description of IBM 7090 Program. The Travelers Research Center, Inc., Hartford, Nov. 1962.

10. Lorenz, E. N., Empirical Orthogonal Functions and Statistical Weather Prediction, 49 pp. Sci. Rpt. 1, Contract AF19(604)-1566, M.I.T., Cambridge, Mass., 1956.

11. Miller, R. G., "The Screening Procedure," pp. 86—95 of Studies in Statistical Weather Prediction, by T. F. Malone. Final Rpt., Contract AF19(604)-1590, The Travelers Research Center, Inc., Hartford, 1959.

12. Reid, J., On the Physical Prediction of Cloud and Fog Formation and the Change in Ceiling Height. Rpt. TRC-16, Contract Cwb-10041, The Travelers Research Center, Inc., Hartford, Aug. 31, 1961.

13. Thomasell, A., Jr., and J. G. Welsh, The Objective Analysis of Sea-level Pressure and Surface Temperature, Dew Point, and Wind. Tech. Rpt. 7044-36, The Travelers Research Center, Inc., Hartford, Nov. 1962.